BAYESIAN SPATIAL MODELLING OF STUNTING CASES IN SOUTH SULAWESI PROVINCE: INFLUENTIAL FACTORS AND RELATIVE RISK

Aswi Aswi[1], Bobby Poerwanto[2], Sudarmin[3] and Nurwan[4]

[1,2,3,4] Statistics Department, Universitas Negeri Makassar, Indonesia
aswi@unm.ac.id

Abstract. Stunting remains a significant public health issue in Indonesia, and numerous research studies have been conducted to address this problem. In 2021, The Bayesian Spatial Conditional Autoregressive (CAR) Localized model was implemented across all 34 Indonesian provinces, revealing that approximately 56% of these provinces are at high risk of stunting. Furthermore, the Bayesian Spatial CAR Leroux model was employed to simulate the relative risk (RR) of stunting cases in one of Indonesia's provinces, South Sulawesi. The main objectives of this study were to determine the most suitable Bayesian CAR Localized model, estimate the RR of stunting, and identify the factors influencing stunting cases in South Sulawesi Province. Data on the number of stunting cases in each district of South Sulawesi Province in 2021 were collected from the South Sulawesi Provincial Health Service and utilized in this study. Population data for 2021 were obtained from the South Sulawesi Provincial Central Statistics Agency. Three covariates were included in this study: the number of people living in poverty, the number of malnourished children, and the number of children with complete basic immunizations. The findings revealed that the Bayesian spatial CAR Localized model with a hyperprior Inverse-Gamma IG (1;0.01) and two clusters, incorporating all three variables, was the most suitable model for predicting stunting cases in South Sulawesi Province in 2021. The number of people living in poverty and the number of malnourished children were positively correlated with the risk of stunting. Conversely, the number of children who have received all their baseline immunizations was inversely associated with the risk of stunting. Stunting affected approximately 54.17% of districts in South Sulawesi Province, with Jeneponto having the highest RR of stunting (RR=1.37) and Makassar having the lowest RR (RR=0.68) among the districts in the province.

Keywords: Bayesian approach, Relative Risk, Spatial CAR Localised, Stunting

1 Introduction

Stunting in children under the age of five is a manifestation of inadequate growth due to prolonged malnutrition [1]. Malnutrition in children who are stunted occurs from the time the baby is in the womb until after birth, which is commonly referred to as the

N. Djam'an et al. (eds.), *Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023)*, Advances in Computer Science Research 109,
https://doi.org/10.2991/978-94-6463-332-0_11

First 1,000 Days of Life (HPK). Stunting is still a major public health problem in developing countries including Indonesia. Addressing stunting has become a priority target both globally and in Indonesia.

Research on modelling stunting cases in Indonesia has been conducted, involving the implementation of the Bayesian Localized spatial Conditional Autoregressive (CAR) model to model stunting cases in all 34 provinces in Indonesia in 2021 [2]. The research findings suggest that around 56% of Indonesian provinces are at a significant risk of experiencing stunting. Sulawesi Barat has the highest Relative Risk (RR) for stunting, with East Nusa Tenggara and West Papua following closely behind. Conversely, Jakarta has the lowest RR for stunting, followed by North Sulawesi and South Sumatra. Additionally, another study has focused on modeling the RR of stunting cases in one Indonesian province, South Sulawesi, using the Bayesian Spatial CAR Leroux model, without incorporating covariates [3]. Another study has explored the factors influencing stunting cases in South Sulawesi Province in 2020, incorporating covariates such as the percentage of poverty, the percentage of exclusive breastfeeding, and the percentage of children aged 0-59 months who were malnourished [4]. Their findings indicate that both the percentage of poverty and the percentage of malnutrition among children aged 0-59 months contribute to an increase of stunting. This research aims to determine the best Bayesian CAR Localised model, estimate the RR of stunting, and identify the factors that influence stunting cases in South Sulawesi Province in 2021.

2 Methods

2.1 Data

The number of stunting cases in each district in South Sulawesi province for the year 2021 was obtained from the South Sulawesi Provincial Health Service. The covariates used include the following variables: the number of people living in poverty (X_1), the number of malnourished children (X_2), and the number of children with complete basic immunizations (X_3).

2.2 Spatial Autocorrelation

Moran's Index (MI) is a commonly used metric for assessing the extent of spatial correlation in both ordinal and interval data. It was introduced by Moran in 1950 [5]. However, Moran's I tends to underestimate spatial autocorrelation in cases involving fewer than 100 areas. To address this limitation, a modified version known as Modified Moran's I (MMI) was introduced [6]. MMI is specifically designed to detect spatial dependence even when dealing with a limited number of areas. The calculation of MMI statistics is performed as follows.

$$\mathrm{MMI} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})\left(\sum_{j=1}^{n} w_{ik}Y_k-\bar{Y}\right)}{\left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]^{1/2}\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ik}Y_k-\bar{Y}\right)^2\right]^{1/2}}$$

where n represent the total number of areas, Y_i and Y_k denote the observed value in the specific areas i and k, $\bar{X}$ represent the average of all the X values across the n areas, ω_{ik}

signifies the spatial weight matrix. The typical choice for a binary spatial matrix in Province the context of areal data analysis is a first-order adjacency weight matrix which is defined [6] as follows:

$$\omega_{ij} = \begin{cases} 1 & \text{if the areas } i \text{ and } j \text{ share a boundary} \\ 0 & \text{otherwise} \end{cases}$$

A comprehensive description of MMI can be found in various scholarly publications [6, 7]. This study has chosen to utilize queen contiguity among the three available spatial adjacency matrix forms, as it has the potential to enhance the model's overall performance.

2.3 Relative Risk

A straightforward way to assess disease risk in regions is by calculating the standardized incidence ratio (SIR), which is the ratio of observed disease cases to the expected cases in a given area. Nevertheless, when dealing with small areas with limited populations or small sample sizes, SIRs can sometimes yield misleading and unreliable results [8]. In such cases, it is preferable to estimate disease risk using Bayesian hierarchical models. Utilizing Bayesian hierarchical models, is more advantageous when assessing the relative risk (RR) in smaller regions than relying solely on unadjusted SIR. Bayesian approaches allow for the integration of information from neighboring areas via prior distributions and the inclusion of covariates within the model, which helps in smoothing out or moderating extreme values, resulting in more robust and dependable risk estimates. An RR of one signifies that the area has a risk level similar to the average of all the areas, whereas RR exceeding one indicate a higher risk than the overall average, and RR below one signify a lower risk [9, 10].

2.4 Model Formulation

The Bayesian spatial CAR localized model [11] was employed to estimate the stunting risk and examine the clusters of stunting cases, both with and without additional factors. This model comprises two essential elements: a spatial random effect (ui) and the clustering components (λ_{Z_i}), which allow for distinct neighborhood random

A frequently employed method for assessing the RR of diseases, known as a Poisson log-linear model [12] was employed to model the number of stunting cases (y_i) as follows:

$$y_i \sim \text{Poisson}(E_i \theta_i) \text{ for } i = 1, 2, 3, \ldots, 24 \text{ areas}$$

$$log(\theta_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_i + \lambda z_i$$

E_i represents the expected cases count, and θ_i is the relative risk in the i^{th} areas. β_0 is the baseline level of RR, while β_1, β_2 and β_3 represent the coefficients for covariate. The spatial structured random effect is modelled using an intrinsic conditional autoregressive (CAR) prior as follows:

$$(u_i | u_k, i \neq k, \tau_u^2) \sim N\left(\frac{\sum_k \mathrm{k}\omega_{ik}}{\sum_k \omega_{ik}}, \frac{\tau_u^2}{\sum_k \omega_{ik}}\right)$$

The spatial weight matrix, denoted ω_{ik}, is constructed using a combination of binary spatial matrix and a first-order adjacency weight matrix. To conduct a sensitivity analysis, five different hyperpriors were applied to the variance component τ_u^2. These hyperpriors include the default choice in CARBayes, which is Inverse-Gamma IG(1; 0.01), as well as four alternative hyperpriors IG (1; 0.1), IG(0.1; 0.1), IG(0.5, ;0.5) and IG(0.5 ;0.0005). Additional insights into the Bayesian spatial CAR localized model are available in particular references [13-16].

We conducted all the analyses in R software, specifically using version 4.2.2 [17] and we employed the CARBayes package version 5.3 [18] to estimate the model parameters. We generated Markov chain Monte Carlo (MCMC) samples by running 12,000 iterations and retaining 13,300 MCMC samples after discarding the initial 1,300 as a burn-in. To assess the convergence of the MCMC, we created trace and density plots. We assessed the suitability of the model formulation and the combination of covariates by using several criteria, including the Deviance Information Criterion (DIC) [19], the Watanabe Akaike Information Criterion (WAIC) [20] and the Modified Moran's I (MMI) [6, 7] for the residuals. Additionally, we considered whether the 95% posterior credible interval contains zero. A lower value of DIC, WAIC, and MMI for residuals indicates a better fit of the model. If necessary, we can provide the R code used in this study upon request.

3 Results and Discussion

3.1 Descriptive Analysis

The total population in South Sulawesi Province in 2021 is 9139531 people with an average of 380813.8, the median 312944.5, and the standard deviation 276480.3. Meanwhile, the total number of stunting cases in South Sulawesi Province was 2535182 cases, with an average of 105632.6, a median of 83526 and a standard deviation of 69685.83. The areas with the largest and the lowest stunting cases are Bone Regency (275,102 cases) and Pare-Pare City (37,925 cases), respectively. Map showing the number of stunting cases in each district in South Sulawesi Province in 2021 is given in Figure 1.

Referring to Figure 1, the data reveals that Bone (ID=2) has the highest number of stunting cases, totaling 275,102 cases, followed by Makassar (ID=10) with 268,392 cases, and Gowa (ID=5) with 255,194. Conversely, Pare-Pare (ID=14) reports the lowest number of stunting cases at 37,925 cases, followed by Selayar (ID=16) with 38,219 cases, and Bantaeng (ID=24) with 44,524 cases. Furthermore, the highest number of populations is Makassar (ID=10) with 1,427,619, followed by Bone (ID=2) with 806,750, and Gowa (ID=5) with 773,315. In contrast, the lowest number of population is Selayar (ID=16) with 137,974, followed by Pare-Pare (ID=14) with 152,922, and Barru (ID=1) with 185,525.

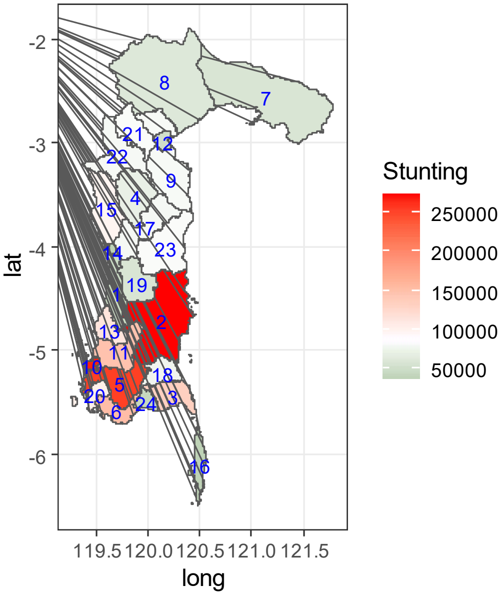

Fig. 1. Map showing the number of stunting cases in each district in South Sulawesi Province in 2021

3.2 Bayesian Spatial CAR Localised Model

Bayesian spatial CAR Localised model with G=2, G=3, dan G=5 and different 5 hyperpriors were used in modeling stunting cases. We have considered different models with various covariate combinations and assessed the convergence of the MCMC. Only combination models with convergent MCMC results are included in this paper. The values of DIC, WAIC and MMI for residual, Credible Interval (CI) for covariate and the number of areas included in the cluster (cluster composition) for CAR Localised model with G=2, G=3, and G=5 are given in Table 1, 2 and 3, respectively.

Table 1 provides the results of the Bayesian spatial CAR Localised model with G=2 with the inclusion of all three covariates. The results indicate that the model using hyperprior IG(1;0.01) stands out with the lowest DIC and WAIC values, labeled as M1. Additionally, M1 has a residual MMI value that is closest to zero. It is worth noting that the number of equivalent regions within each cluster (G1 and G2) remains consistent across different hyperpriors. The covariate of number of people living in poverty and the number of malnourished children were positively correlated with the risk of stunting. However, the number of children who have received all their baseline immunizations has a negative and significant associated with the risk of stunting.
Summary Statistics for Bayesian Spatial CAR localised models with 3 clusters, including DIC, WAIC, MMI for residuals, CI, and cluster composition were given in Table 2.

Table 1. DIC, WAIC, MMI for residuals, CI and cluster composition for CAR Localised model with G=2.

Hyperpriors	Model	Covariate	DIC	WAIC	MMI residual	Credible Interval		(Cluster Composition)	
						2.5%	97.5%	G1	G2
IG (1;0.01)	M1	X_1	**368.25**	**362.56**	**-0.12**	0.014	0.017	11	13
		X_2				0.042	0.046		
		X_3				-0.150	-0.146		
IG (1;0.1)	M2	X_1	368.66	362.93	0.15	0.014	0.015	11	13
		X_2				0.041	0.043		
		X_3				-0.146	-0.144		
IG (0.1;0.1)	M3	X_1	368.76	363.38	-0.22	0.013	0.014	11	13
		X_2				0.040	0.044		
		X_3				-0.148	-0.144		
IG (0.5;0.5)	M4	X_1	369.41	364.45	-0.43	0.012	0.015	11	13
		X_2				0.041	0.044		
		X_3				-0.148	-0.145		
IG (0.5;0.0005)	M5	X_1	368.56	362.61	-0.73	0.014	0.017	11	13
		X_2				0.042	0.045		
		X_3				-0.149	-0.146		

Table 2. DIC, WAIC, MMI for residuals, CI and cluster composition for CAR Localised model with G=3.

Hyperpriors	Model	Covariate	DIC	WAIC	MMI residual	Credible Interval		Cluster Composition		
						2.5%	97.5%	G1	G2	G3
IG (1;0.01)	M6	X_1	**369.50**	**366.61**	-0.62	0.014	0.015	9	11	4
		X_2				0.042	0.044			
		X_3				-0.148	-0.146			
IG (1;0.1)	M7	X_1	370.76	368.11	**-0.30**	0.015	0.018	9	11	4
		X_2				0.040	0.043			
		X_3				-0.148	-0.145			
IG (0.1;0.1)	M8	X_1	371.59	370.27	-0.48	0.012	0.015	9	11	4
		X_2				0.038	0.043			
		X_3				-0.147	-0.142			
IG (0.5;0.5)	M9	X_1	370.45	368.08	-0.44	0.014	0.018	9	11	4
		X_2				0.039	0.042			
		X_3				-0.146	-0.144			
IG (0.5;0.0005)	M10	X_1	370.75	368.15	-0.57	0.014	0.015	9	11	4
		X_2				0.043	0.045			
		X_3				-0.148	-0.147			

Table 2 indicates that the model with the hyperprior IG(1;0.01) (M6) has the lowest DIC and WAIC values. Additionally, the model with the hyperprior IG(1;0.1) (M7) exhibits a residual MMI value closest to zero. Although the DIC and WAIC values for

M6 and M7 are relatively similar, the MMI residual value for M7 is smaller than that of M6. Consequently, the M7 model is more suitable than M6. The number of regions in each cluster (G1, G2, and G3) remains the same across different hyperpriors. It is worth noting that covariates such as the proportion of the poor population and poor nutrition show a positive and significant relationship with the risk of stunting, while complete basic immunization exhibits a negative and significant relationship with the risk of stunting.

Summary statistics for Bayesian Spatial CAR Localised Models with 5 clusters, including DIC, WAIC, MMI for residuals, CI, and cluster composition were given in Table 3. Table 3 indicates that the model with the hyperprior IG(0.5;0.0005) (M15) has the lowest DIC and WAIC values. Additionally, the model with the hyperprior IG(1;0.1) (M12) exhibits a residual MMI value closest to zero. Although the DIC and WAIC values for M15 and M12 are relatively similar, the MMI residual value for M12 is smaller than that of M15. Consequently, the M12 model is more suitable than M15. The number of regions in each cluster (G1, G2, G3, G4, and G5) remains the same across different hyperpriors. It is worth noting that covariates such as the proportion of the poor population and poor nutrition show a positive and significant relationship with the risk of stunting, while complete basic immunization exhibits a negative and significant relationship with the risk of stunting.

Table 3. DIC, WAIC, MMI for residuals, CI and cluster composition for CAR Localised model with G=5.

Hyperpriors	Model	Co-variate	DIC	WAIC	MMI residual	Credible Interval		Cluster Composition				
						2.5%	97.5%	G1	G2	G3	G4	G5
IG (1;0.01)	M11	X_1	373.84	378.32	-0.73	0.014	0.015	2	4	5	9	4
		X_2				0.041	0.043					
		X_3				-0.148	-0.146					
IG (1;0.1)	M12	X_1	371.90	374.79	**-0.24**	0.013	0.014	2	4	5	9	4
		X_2				0.042	0.043					
		X_3				-0.146	-0.146					
IG (0.1;0.1)	M13	X_1	374.22	380.00	-0.67	0.010	0.014	2	4	5	9	4
		X_2				0.042	0.045					
		X_3				-0.148	-0.146					
IG (0.5;0.5)	M14	X_1	371.92	374.41	-0.28	0.014	0.017	2	4	5	9	4
		X_2				0.042	0.045					
		X_3				-0.149	-0.146					
5;0.0005)	M15	X_1	**371.08**	**373.50**	-0.82	0.014	0.016	2	4	5	9	4
		X_2				0.037	0.042					
		X_3				0.146	0.141					

Based on all the model selection criteria used in this study, the Bayesian Spatial Localized CAR model with a hyperprior of IG(1;0.01) and G=2 (M1) is the preferred choice for estimating the relative risk of stunting in South Sulawesi province. Based on the best model, it is concluded that the number of people living in poverty (X_1) and the number of malnourished children (X_2) are positively correlated with the risk of stunting, while the number of children with complete basic immunizations (X_3) is negatively correlated with the risk of stunting. The estimated RR values for each district using the preferred model can be found in Table 4 and Figure 2.

Table 4. The estimated RR values for each district using the preferred model

ID	Districs	RR	ID	Districst	RR
1	Barru	0.95	13	Pangkep	1.19
2	Bone	1.23	14	Pare-Pare	0.89
3	Bulukumba	1.11	15	Pinrang	0.88
4	Enrekang	1.15	16	Selayar	1.00
5	Gowa	1.19	17	Sidrap	0.92
6	Jeneponto	1.37	18	Sinjai	1.09
7	Luwu Timur	0.72	19	Soppeng	0.92
8	Luwu Utara	0.70	20	Takalar	1.25
9	Luwu	0.82	21	Toraja Utara	1.18
10	Makassar	0.68	22	Toraja	1.05
11	Maros	1.35	23	Wajo	0.81
12	Palopo	1.03	24	Bantaeng	0.81

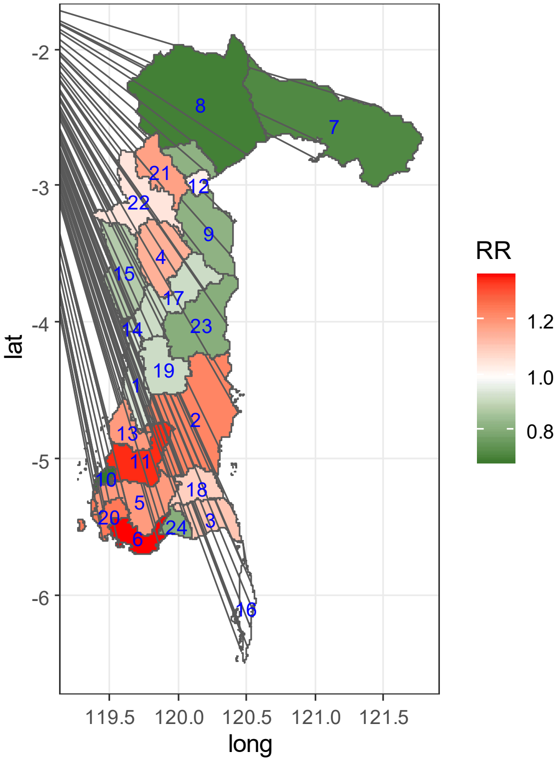

Fig 2. Thematic Map of RR using Bayesian spasial CAR Localized Model with G=2

According to Table 4 and Figure 2, Jeneponto district has the highest RR, followed by Maros district and Takalar. Conversely, Makassar City has the lowest RR, followed by North Luwu district and East Luwu. Approximately 54.17% of districts within South Sulawesi Province experienced the impact of stunting.

Conclusions

The preferred model for assessing the RR of stunting in South Sulawesi province in 2021 is the Bayesian Spatial CAR model with a hyperprior IG(1;0.01) with G=2, which incorporates all three variables. There was a positive correlation between the number of people living in poverty and the number of malnourished children with the risk of stunting. Conversely, the risk of stunting was inversely associated with the number of children who had received all their baseline immunizations. Stunting was observed in approximately 54.17% of districts in South Sulawesi Province, with Jeneponto having the highest RR of stunting (RR=1.37) and Makassar having the lowest RR (RR=0.68) among the districts.

Acknowledgement

The authors wish to acknowledge the Universitas Negeri Makassar. This work was funded by DIPA Universitas Negeri Makassar Number: SP DIPA 023.17.2.677523/2023, February 15, 2023. according to the Decree of the Rector of Universitas Negeri Makassar Number: 299/UN36/HK/2023 on March 05, 2023.

References

1. Statistik BP. Profil Kesehatan Ibu dan Anak 2022. Jakarta: Badan Pusat Statistik; 2022. 405 p.
2. Azis AA, Aswi A. Spatial clustering of stunting cases in Indonesia: A Bayesian approach. Communications in Mathematical Biology and Neuroscience. 2023;2023:1-11.
3. Aswi A, Sukarna S, Nurhilaliyah. Pemetaan Risiko Relatif Kasus Stunting di Provinsi Sulawesi Selatan. Sainsmat: Jurnal Ilmiah Ilmu Pengetahuan Alam. 2022;11(1):11-20.
4. Aswi AS, Sukarna. Pemodelan Spasial Bayesian dalam Menentukan Faktor yang Mempengaruhi Kejadian Stunting di Provinsi Sulawesi Selatan. Journal of Mathematics, Computations, and Statistics. 2022;5(1):1-11.
5. Moran PAP. Notes on continuous stochastic phenomena. Biometrika. 1950;37(1-2):17.
6. Carrijo TB, Da Silva AR. Modified Moran's I for Small Samples. Geogr Anal. 2017;49(4):451-67.
7. Aswi A, Cramb S, Duncan E, Mengersen K. Detecting Spatial Autocorrelation for a Small Number of Areas: a practical example. Journal of physics Conference series. 2021;1899(1):12098.
8. Moraga P. Geospatial health data : modeling and visualization with R-INLA and Shiny. Boca Raton: CRC Press; 2020.
9. Banerjee S, Carlin BP, Gelfand AE. Hierarchical modeling and analysis for spatial data: Crc Press; 2014.

10. Alston CL, Mengersen KL, Pettitt AN. Case Studies in Bayesian Statistical Modelling and Analysis: wiley; 2012. 1-464 p.
11. Lee D, Sarran C. Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. Environmetrics. 2015;26(7):477-87.
12. Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. Epidemiol Infect. 2019;147.
13. Lee D, Lawson A. Quantifying the Spatial Inequality and Temporal Trends in Maternal Smoking Rates in Glasgow. Ann Appl Stat. 2016;10(3):1427-46.
14. Aswi A, Cramb S, Duncan E, Mengersen K. Evaluating the impact of a small number of areas on spatial estimation. Int J Health Geogr. 2020;19(1):39-.
15. Aswi A, Mauliyana A, Tiro MA, Bustan MN. Relative Risk of Coronavirus Disease (Covid-19) in South Sulawesi Province, Indonesia: Bayesian Spatial Modeling. Media Statistika. 2022;14(2):158-69.
16. Aswi A, Tiro MA, Sudarmin S, Sukarna S, Cramb S. The Interplay Between Clusters, Covariates, and Spatial Priors in Spatial Modelling of Covid-19 in South Sulawesi Province, Indonesia. Media Statistika. 2022;15(1):48-59.
17. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
18. Lee D. CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. J Stat Softw. 2013;55(13):1-24.
19. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B, Statistical methodology. 2002;64(4):583-639.
20. Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. J Mach Learn Res. 2010;11:3571-94.